ADVANCES IN APPLIED MATHEMATICS 4, 103–123 (1983)

Discrete-Time Feedback Systems: An Operator Theoretic Approach

AVRAHAM FEINTUCH

Department of Mathematics, Ben Gurion University, Beer Sheva, Israel

INTRODUCTION

In a recent paper a survey of the mathematical foundations of the analysis and processing of time-ordered data was given from an operator-theoretic point of view [9]. The main operator-theoretic tools that appeared were the spectral theorem for unitary operators and the Wold decomposition theorem for isometric operators.

In this paper we will attempt to do the same for the theory of linear discrete time feedback systems. Recent results in operator theory and systems theory will allow us to present quite a general, broad theory which will cover such topics as stability, stabilization, estimation and filtering. The major advantages of this theory is that it doesn't require any hypothesis of time-invariance nor do the conceptual problems become any more difficult in the multi-channel case. In fact, these notions play no role in the theory. The main operator theoretic tool that is used is an abstract inner-outer factorization theorem due independently to Arveson [1] and Rissanen [18]. The mathematical framework is that of Hilbert Resolution Space introduced by Porter [13] and Saeks [14]. Few of the results mentioned here are new. However, the essential completeness of the discrete time theory does not seem to have been noticed before. It should be mentioned that an equally complete theory for continuous time systems is not possible. This is because Arveson's factorization theorem and the rather simple invertibility criteria do not extend to the continuous time case ([11], [3]).

The contents of this paper are organized as follows:

Section 1: Hilbert Resolution Space and Extended Space

Section 2: Arveson's Theorem and its consequences

Section 3: Feedback Systems: basic concepts

Section 4: Least Squares Theory

Section 5: Stochastic Control

0196-8858/83 $7.50

1. Hilbert Resolution Space and Extended Space

Let $\mathcal{H}$ be a separable complex Hilbert Space, and let $\mathcal{E}$ be a totally ordered family of orthogonal projections on $\mathcal{H}$ which:

(i) is order isomorphic to (a subset of) the integers Z.

(ii) contains O, I.

(iii) is closed under the lattice operations $\cap$, $\vee$ (closed linear span).

The pair $(\mathcal{H}, \mathcal{E})$ is a discrete time Hilbert Resolution Space. The best known example is, of course, the sequence space $l^2(-\infty, \infty;)$ of square summable sequences of vectors from a Hilbert space.

Throughout this paper we will model linear systems by linear operators acting on spaces and satisfying the physical property of causality. This is the "input-output" point of view and we will avoid introducing the notion of state-space. While a fairly complete abstract state-space theory exists in this framework [17] and plays an important role in various filtering problems [4], we will avoid this area here.

Causality can be defined in terms of the pair $(\mathcal{H}, \mathcal{E})$ as follows [5–13].

Definition 1.1. The operator T on $(\mathcal{H}, \mathcal{E})$ is causal if, for each $n \in Z$,

$$E^n T = E^n T E^n.$$

If we take $\mathcal{H} = l^2(-\infty, \infty)$, and $\mathcal{E}$ to be the family of projections defined by

$$E^n\{a_k\} = \{b_k\},$$

where

$$\begin{aligned} b_k &= a_k, & k &\leqslant n \\ &= 0, & k &> n, \end{aligned}$$

then the causal bounded linear operators are simply those that have lower triangular representations with respect to the standard orthonormal basis on $\mathcal{H}$.

Most notions related to stability can not be discussed on $\mathcal{H}$, because one must allow inputs to the system which aren't vectors on the Hilbert space. Classically this, technically, was taken care of by considering the extended space of the Hilbert space under consideration. For example, if $\mathcal{H}$ was $l^2(-\infty, \infty)$ one defined the extended space $l_e^2(-\infty, \infty)$ to be the family of sequences which when truncated at some point n were square summable; i.e. $l_e^2(-\infty, \infty) = \{\{a_k\}_{-\infty}^{\infty}: \Sigma_{k=-\infty}^{n} |a_k|^2 < \infty \text{ for all } n \in Z\}$. We will construct an abstract extended space for the pair $(\mathcal{H}, \mathcal{E})$. While this construction works equally well for arbitrary time sets, we will restrict ourselves here and throughout to the discrete time case [6].

For each $n \in Z$, define the seminorm

$$\|h\|^n = \|E^n h\|, \qquad h \in \mathcal{H},$$

If $h \neq 0$, then $E^n h \neq 0$ for some $n \in Z$ and therefore, for this n, $\|h\|^n \neq 0$. Thus the family of seminorms, $\| \quad \|^n$, separate $\mathcal{H}$ and defines a Hausdorff topology τ on $\mathcal{H}$ ([10]). In fact, using a standard result on topological vector spaces, we have ([10], p. 114):

THEOREM 1.2. *The map $h \to |h|$ from $\mathcal{H}$ into R_+, where*

$$|h| = \sum_{n=-\infty}^{\infty} \frac{1}{2^n} \frac{\|h\|^n}{1 + \|h\|^n},$$

defines a translation invariant metric on $\mathcal{H}$ with the following properties:

(a) $|h| = 0$ *if and only if* $h = 0$,

(b) $|h| = |-h|$,

(c) $|h + g| \leqslant |h| + |g|$,

(d) $|\lambda| \leqslant 1$ *implies* $|\lambda h| \leqslant |h|$,

(e) $\lambda \to 0$ *implies* $|\lambda h| \to 0$ *for every* $h \in \mathcal{H}$.

Furthermore, the topology defined by the metric $\delta(h, g) = |h - g|$ is just the topology τ defined by the family of seminorms $\{\| \quad \|^n : n \in Z\}$.

Clearly, the topology τ on $\mathcal{H}$ is weaker than the norm topology, for if $\langle h_i \rangle \in \mathcal{H}$ and $\|h_i - h\| \to 0$, then $\|E^n(h_i - h)\| \leqslant \|h_i - h\| \to 0$ for all $n \in Z$. We will call this topology the resolution topology on $\mathcal{H}$.

Which operators continuous with respect to the norm topology on $\mathcal{H}$ remain continuous with respect to the resolution topology? The operators which represent physical systems, namely the causal ones, retain their continuity [6].

THEOREM 1.3. *Every causal bounded operator on $\mathcal{H}$ is continuous in the resolution topology.*

Proof. By translation invariance, it is enough to consider continuity at 0. So suppose $\langle h_i \rangle \in \mathcal{H}$ such that $|h_i| \to 0$. Then for all n, $\|h_i\|^n \to 0$.

If T is causal and bounded, then

$$\begin{aligned}\|Th_i\|^n &= \|E^n T h_i\| = \|E^n T E^n h_i\| \\ &\leqslant \|E^n\| \|T\| \|E^n h_i\| \\ &\leqslant \|T\| \|h_i\|^n \to 0.\end{aligned}$$

Thus T is continuous.

It is easy to see that with respect to the metric δ on $\mathcal{H}$, $\mathcal{H}$ is not complete. Let $\mathcal{H}_e$ denote the completion of $\mathcal{H}$ in the metric δ. Since each E^n is causal, it is continuous in the resolution topology and therefore has a unique continuous extension E_e^n to $\mathcal{H}_e$.

DEFINITION 1.4. The pair $(\mathcal{H}_e, \mathcal{E}_e)$ is the extended resolution space of $(\mathcal{H}, \mathcal{E})$.

The relationship between $\mathcal{H}$ and $\mathcal{H}_e$ is summarized in the following proposition [6].

PROPOSITION 1.5. *Let $(\mathcal{H}_e, \mathcal{E}_e)$ be the extended space for $(\mathcal{H}, \mathcal{E})$. Then*:

(i) $E_e^n h_e \in \mathcal{H}$ *for all* $h_e \in \mathcal{H}_e$ *and* $n \in Z$;

(ii) *if* $\|\ \|_e^n$ *denotes the extension of* $\|\ \|^n$ *to* $\mathcal{H}_e$, *then*

$$\|h_e\|_e^n = \|E_e^n h_e\| \text{ for all } h_e \in \mathcal{H}_e \text{ and } n \in Z;$$

(iii) *if* $h_e \in \mathcal{H}_e$, *then* $h_e \in \mathcal{H}$ *if and only if* $\sup_n \|h_e\|_e^n < \infty$.

We note that the concept of causality generalizes naturally to $\mathcal{H}_e$ via the equality

$$E_e^n T = E_e^n T E_e^n, \qquad n \in Z,$$

for operators on $\mathcal{H}_e$. If A is a continuous causal operator on $\mathcal{H}$ and A_e is its unique continuous extension to $\mathcal{H}_e$, then it is clear that A_e is causal. Also, by the proof of Theorem 1.3, and Proposition 1.5

$$\|A_e h_e\|_e^n \leqslant \|A\| \|h_e\|_e^n$$

for all $h_e \in \mathcal{H}_e$ and $n \in Z$. This motivates the definition of stability.

DEFINITION 1.6. An operator T on $\mathcal{H}_e$ is *stable* if there exists $0 < M < \infty$ such that

$$\|Th_e\|_e^n \leqslant M \|h_e\|_e^n$$

for all $n \in Z$, $h_e \in \mathcal{H}_e$.

Remark 1.7. Stability is much stronger than continuity. Since $\mathcal{H}_e$ is a topological vector space which is not a normed space, the notions of boundedness and continuity are not equivalent.

All stable operators on $\mathcal{H}_e$ are characterized in the next result [6].

THEOREM 1.8. *Suppose T is a linear operator on $\mathcal{H}_e$. Then T is stable if and only if $T = S_e$ for some bounded causal operator S on $\mathcal{H}$.*

Proof. Suppose T is stable, and consider the operator $E_e^n T[I - E_e^n]$ on $\mathcal{H}_e$. Then $\|E_e^n T[1 - E_e^n] h_e\|_e^n = \|E_e^n (E_e^n T[I - E_e^n]) h_e\|$ by proposition 1.5

(ii), and since $E_e^n E_e^n = E_e^n$ this is just $\|E_e^n T[I - E_e^n]h_e\|$. By the same proposition this equals

$$\begin{aligned} \|T[I - E_e^n]h_e\|_e^n &\leqslant M\|[I - E_e^n]h_e\|^n \\ &= M\|E^n[I - E_e^n]h_e\| \\ &= 0 \end{aligned}$$

for each $h_e \in H_e$. Thus $E_e^n T = E_e^n T E_e^n$, and T is causal on $(\mathcal{H}_e, \mathcal{E}_e)$. If $h \in \mathcal{H}$, then

$$\begin{aligned} \sup_n \|Th\|_e^n &\leqslant \sup_n [M\|h\|_e^n] \\ &= \sup_n [M\|h\|^n] \leqslant M\|h\| \end{aligned}$$

Thus $Th \in \mathcal{H}$ by (iii) of Proposition 1.5. Also

$$\begin{aligned} \|Th\| = \lim_n \|E^n Th\| &= \lim \|Th\|^n \\ &\leqslant \sup_n \|Th\|^n = \sup_n \|Th\|_e^n \\ &\leqslant M\|h\|. \end{aligned}$$

Thus the restriction of T to $\mathcal{H}$ is bounded. The only if part of the theorem follows from the paragraph preceding the theorem.

2. The Arveson-Rissanen Theorem and Its Consequences

Let $\{\mathfrak{M}_n: n \in Z\}$ be a family of closed subspaces of $\mathcal{H}$ satisfying $\mathfrak{M}_n \subseteq \mathfrak{M}_{n+1}$, $\cap_n \mathfrak{M}_n = \{0\}$, $\vee_n \mathfrak{M}_n = \mathcal{H}$. Let $\mathfrak{R} = \{A \in \mathfrak{B}(\mathcal{H}): A\mathfrak{M}_n \subseteq \mathfrak{M}_n$ for all $n \in Z\}$. It is easily seen that $\mathfrak{R}$ is a (weakly closed) Banach algebra containing the identity.

The main result of this section is a special case of a result of Arveson [1].

THEOREM 2.1. *Suppose T is a invertible operator on $\mathcal{H}$. Then $T = UA$, where U is unitary and $A, A^{-1} \in \mathfrak{R}$.*

Before giving the proof, we consider the simplest case. This will make the idea of the proof quite transparent.

Suppose $\mathcal{H} = l^2(0, \infty)$, and $\{e_i\}_{i=0}^{\infty}$ is the standard basis on $\mathcal{H}$. Let $\mathcal{E}$ be the family of truncation projections $\{P_n\}_{n=0}^{\infty}$, and $\mathfrak{M}_n = \vee_{i=0}^{n} e_i$. Suppose T is an invertible operator. Then $\{Te_0, Te_1, \ldots Te_n\}$ is a basis for $T\mathfrak{M}_n$. If we apply the Gram–Schmidt process to the set $\{T_{ei}\}_{i=0}^{\infty}$, we obtain an orthonormal basis $\{f_i\}$ of $\mathcal{H}$ such that $\vee_{i=0}^{n} f_i$ is an orthonormal basis for $T\mathfrak{M}_n$. Let U be the unitary operator which maps e_i to f_i. Then let $A = U^*T$. Clearly $U = TA$, and since $Ae_i = U^*Te_i = U^*(\Sigma_{j=0}^{i}\alpha_j f_j) = \Sigma_{j=0}^{i}\alpha_j e_j$, we have $A\mathfrak{M}_n \subset \mathfrak{M}_n$. Since A is invertible, $A^{-1}\mathfrak{M}_n \subset \mathfrak{M}_n$ and $T = UA$ is the required factorization. The proof of the general case is based on the same idea.

Proof. Let $\mathfrak{L}_n = T\mathfrak{M}_n$ for each n. Since T is invertible, $\cap\{\mathfrak{L}_n: n \in Z\} = \{0\}$ and $\vee\{\mathfrak{L}_n: n \in Z\} = \mathcal{H}$. Also $T\mathfrak{M}_n \subseteq T\mathfrak{M}_{n+1}$ for all $n \in Z$ by continuity. Thus $\mathfrak{L}_n$ retains the properties of $\mathfrak{M}_n$. Also since dimension is preserved by invertible operators, it is easily seen that $\dim(\mathfrak{M}_n \ominus \mathfrak{M}_{n-1}) = \dim(\mathfrak{L}_n \ominus \mathfrak{L}_{n-1})$ for all $n \in Z$. Since $\mathcal{H} = \Sigma \oplus (\mathfrak{M}_n \ominus \mathfrak{M}_{n-1}) = \Sigma \oplus (\mathfrak{L}_n \ominus \mathfrak{L}_{n-1})$, let U^* be a unitary operator which maps, for each $n \in Z$, $\mathfrak{L}_n \ominus \mathfrak{L}_{n-1}$ isometrically onto $\mathfrak{M}_n \ominus \mathfrak{M}_{n-1}$, and let $A = U^*T$. Then for each n,

$$\begin{aligned} A\mathfrak{M}_n &= U^*T\mathfrak{M}_n = U^*\mathfrak{L}_n \\ &= U^*\Big(\sum_{k\leqslant n} \oplus [\mathfrak{L}_k \ominus \mathfrak{L}_{k-1}]\Big) = \sum_{k\leqslant n} \oplus [\mathfrak{M}_k \ominus \mathfrak{M}_{k-1}] \\ &= \mathfrak{M}_n. \end{aligned}$$

Thus, $A \in \mathfrak{R} \cap \mathfrak{R}^{-1}$ and $T = UA$ gives the desired factorization.

An immediate consequence is the following spectral factorization theorem.

COROLLARY 2.2. *Every invertible positive operator on $\mathcal{H}$ can be factored in the form A^*A with $A \in \mathfrak{R} \cap \mathfrak{R}^{-1}$.*

Proof. Suppose H is an invertible positive operator on $\mathcal{H}$, and let $T = H^{1/2}$, its positive invertible square root. By Theorem 2.1, $T = UA$ with U unitary, and $A \in \mathfrak{R} \cap \mathfrak{R}^{-1}$. Then

$$H = (H^{1/2})^*H^{1/2} = A^*U^*UA = A^*A$$

gives the required factorization.

Remark 1. The relationship between the spectral factorization given here and the classical factorizations arising, for example, in Wiener–Hopf equations has been discussed in detail in [16]. It is clear that the fact that $\{\mathfrak{M}_n\}$ is order-isomorphic to the integers played a fundamental role in the proof of Theorem 2.1. The result is false [11] if the chain of subspaces is order isomorphic to the interval [0, 1]. If, however, it is not required that $A \in \mathfrak{R}^{-1}$, then Theorem 2.1 can be extended to the continuous case [3].

Remark 2. If $M_n = E_n\mathcal{H}$, then $\mathfrak{R}$ is the algebra of causal operators on $\mathcal{H}$. The operators in $\mathfrak{R} \cap \mathfrak{R}^{-1}$ are called the causally invertible operators and play a major role in stability theory as we will see later.

Remark 3. In [1], Theorem 2.1 was proved in a more general form than that given here. We can drop the assumption that T is invertible and simply assume that $\cap_n [TM_n] = 0$. Then U will be an isometry and A an "outer" operator. Thus Theorem 2.1 can be seen as a generalized inner–outer factorization theorem. In fact, if $\mathcal{H} = H^2$ and T is a time invariant (equivalently, analytic Toeplitz) operator with symbol f, then Arveson has shown that the two factorizations are equivalent.

There is also a structural relationship between this theorem and the wold decomposition. In fact, if T is an isometry, then the decomposition of $\mathcal{H}$ in the proof of Theorem 2.1 is just the wold decomposition.

Remark 4. Corollary 2.2 can be generalized to some positive definite symmetric operators which are not necessarily bounded ([19]). If Q is such an operator, with domain $\mathfrak{D}(Q)$ there exists an "outer" operator with $\mathfrak{D}(A) = \mathfrak{D}(Q)$, such that $Q = A^*A$.

We have used the term "outer" in Remarks 3 and 4. While the definition given in [19] is less stringent than that in [1], for invertible operators they are equivalent and this is the case which will be of interest to us. We will thus use the definition given by Rissanen.

DEFINITION 2.3 ([19]). An operator A on $\mathcal{H}$ is outer if (i) $AM_n \subseteq M_n \forall n \in Z$, (ii) $[AM_n] \ominus [AM_{n-1}] \subset M_n \ominus M_{n-1}$ for all $n \in Z$. Here $[AM_n]$ denotes the closure of AM_n.

Note that if $A \in \mathfrak{R}$ is invertible with $A^{-1} \in \mathfrak{R}$, then A is outer.

THEOREM 2.4. *Suppose $A \in \mathfrak{R}$ is an invertible operator. If for all $n \in Z$, $\dim M_n \ominus M_{n-1} < \infty$, then $A^{-1} \in \mathfrak{R}$ if and only if A is outer.*

Proof. We must show that if $A \in \mathfrak{R}$ is outer, then $A^{-1} \in \mathfrak{R}$, or equivalently, that $AM_n = M_n$ for all $n \in Z$. Since A is outer, $AM_n \ominus AM_{n-1} \subset M_n \ominus M_{n-1}$. As in Theorem 2.1, A invertible implies that $\dim(AM_n \ominus AM_{n-1}) = \dim(M_n \ominus M_{n-1})$. Thus $\dim(M_n \ominus M_{n-1}) < \infty$ implies $AM_n \ominus AM_{n-1} = M_n \ominus M_{n-1}$. Since $AM_{n-1} \subseteq AM_n$ we can write AM_n as $AM_n = (AM_n \ominus AM_{n-1}) \oplus AM_{n-1} = (M_n \ominus M_{n-1}) \oplus AM_{n-1}$. In particular $M_n \ominus M_{n-1} \subset AM_n$. Also for $i < n$, $M_i \ominus M_{i-1} \subset AM_i \subset AM_n$. Since, as in Theorem 2.1, $M_n = \Sigma_{i \leqslant n}[M_i \ominus M_{i-1}]$, $M_n \subset AM_n$, and therefore $AM_n = M_n$.

Remark. If $\mathcal{H} = \mathcal{H}^2$ and $M_n = \vee_{i=n}^{\infty} e_i$ and A is time invariant with symbol f, then A, $A^{-1} \in \mathfrak{R}$ if and only if f is an invertible outer function in the classical sense. Thus Theorem 2.4 can be seen as a time-varying analog of this classical result.

3. Feedback Systems, Basic Concepts

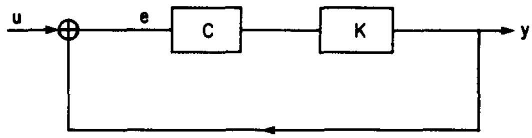
Definition 3.1. Let $(\mathcal{H}_e, \mathcal{S}_e)$ be the extended space of $(\mathcal{H}, \mathcal{S})$. A *basic feedback system* on $(\mathcal{H}_e, \mathcal{S}_e)$ is a pair (K, C) of causal operators on $\mathcal{H}_e$ related by the equations

$$y = KCe,$$

$$e = u + y,$$

with $y, u, e \in \mathcal{H}_e$.

A basic feedback system is characterized by the block diagram below.



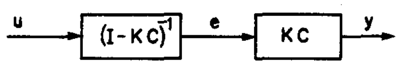
Conceptually, K is the plant and C is a compensator or controller, u is the system input, y is the output, and e is the error or compensator input. There are variations on the basic feedback system which can be considered. These may involve additional inputs and (or) operators. The basic structure and problem will remain unchanged.

The same input–output relationships can be expressed without feedback by the equations

$$e = (I - KC)^{-1}u,$$

$$y = KC(I - KC)^{-1}u,$$

provided $(I - KC)^{-1}$ exists in some sense. The system can then be represented by the block diagram. The first representation will be referred to as the closed loop representation of the system, while the second is the open-loop representation.



Definition 3.2. The basic feedback system (K, C) on $\mathcal{H}_e$ is well-posed if $(I - KC)$ is one-to-one and onto on $\mathcal{H}_e$.

Clearly well-posed feedback systems allow an open loop representation and we will from now on restrict ourselves to well-posed basic feedback systems.

EXAMPLE 3.3. Let $\mathcal{H} = l^2(0, \infty)$ with the usual resolution of the identity. Suppose $y = \{y_i\} \in \mathcal{H}$ and KC is defined by

$$(KCy)_i = \sum_{n=0}^{i-1} g_{in} y_n, \qquad i \geqslant 0,$$

where $\{g_{in} | n < i, \; i = 0, 1, 2, \dots\}$ is a sequence of complex numbers with $g_{in} \neq 0$ for all i, n. Then the equation

$$z = (I - KC)y$$

gives

$$z_i = y_i - \sum_{n=0}^{i-1} g_{in} y_n, \qquad i \geqslant 0.$$

Thus, given $z = \{z_i\}$, one can compute recursively via

$$y_0 = z_0,$$
$$y_i = z_i + \sum_{n=0}^{i-1} g_{in} y_n.$$

In general, $\{y_i\} \in \mathcal{H}_e$, but is not square summable. Thus $(I - KC)^{-1}$ exists as an operator on $\mathcal{H}_e$. Thus the system described above is well posed. We note that $(I - KC)^{-1}$ is causal on $\mathcal{H}_e$, but is not bounded.

Well-posedness is essentially a modeling property of a system. It expresses that a mathematical model is, at least in principle, adequate as a description of a physical system. The definition which follows is that of stability which is a desired property of a feedback system.

DEFINITION 3.4. Let (K, C) be a basic feedback system on $(\mathcal{H}_e, \mathcal{E})$. The system is stable if the operators $(I - KC)^{-1}$ and $KC(I - KC)^{-1}$ are stable operators.

Thus a system is stable if the operators appearing in the corresponding open loop system are stable, or, equivalently, given an input u in $\mathcal{H}$, the error e and output y are also in $\mathcal{H}$.

An important case is the case when K and C are stable operators. The stability of the system (K, C) is equivalent to stability of $(I - KC)^{-1}$. In this case, K and C are extensions of causal bounded linear operators K_1, C_1, respectively, on $\mathcal{H}$. The well posedness of the system and the stability of K and C imply that $(I - K_1C_1)$ is invertible on H. Thus the stability of (K, C) is equivalent to the fact that $(I - K_1C_1)^{-1}$ is a causal operator on $\mathcal{H}$.

This motivates us to consider the well-known "Causal Invertibility Problem." If T is a causal operator invertible on $\mathcal{H}$, when is T^{-1} causal? That this is not always the case even in the discrete time case is seen immediately from the following example.

EXAMPLE 3.5. Let $\mathcal{H} = l^2(-\infty, \infty)$, and $\mathcal{E}$ the family of truncation projections. Let W be the bilateral shift $We_n = e_{n+1}$ defined on the standard orthonormal basis $\{e_i\}_{i=-\infty}^{\infty}$. Then W is causal, but its inverse $W^{-1} = W^*$ is anti-causal.

The answer to this question for the case $\dim(E_n - E_{n-1})\mathcal{H} < \infty$ for all n is given by Theorem 2.4. The causally invertible operators are the invertible causal operators which are outer.

4. Least-Squares Theory

In this section we investigate the formulation and solution of some of the more commonly encountered deterministic optimal control problems. We begin by formulating a general theorem due to Porter [12] which will contain as special cases the servomechanism and regulator problems we will consider. The treatment here and in the following sections strongly follows [5].

THEOREM 4.1. *Let $A \in \mathcal{B}(\mathcal{H})$ be such that $P = A^*A$ is positive definite. Let C^*C be the spectral factorization of P, and let $w \in E_n\mathcal{H}$, $z \in \mathcal{H}$. Then the cost functional*

$$J(B) = \|z - ABw\|^2$$

is minimized over $B \in \mathcal{C}$ by any $B_0 \in \mathcal{C}$ such that $B_0 w = C^{-1}E_nC^{*-1}A^*z$.

Proof. Since $w \in \mathcal{H}_n = E_n\mathcal{H}$, for any $B \in \mathcal{C}$ we have

$$CBw = CBE_nw = E_nCBE_nw = E_nCBw.$$

Thus

$$\begin{aligned}(z, ABw) &= (A^*z, Bw) = (C^*C^{*-1}A^*z, Bw)\\ &= (C^{*-1}A^*z, CBw)\\ &= \left(C^{*-1}A^*z, E_nCBw\right)\\ &= \left(E_nC^{*-1}A^*z, CBw\right)\\ &= \left(C^{*-1}C^*CC^{-1}E_nC^{*-1}A^*z, CBw\right)\\ &= \left(C^{*-1}A^*AB_0w, CBw\right)\\ &= \left(AB_0w, AC^{-1}CBw\right) = (AB_0w, ABw).\end{aligned}$$

A similar argument shows that

$$(z, AB_0w) = (AB_0w, AB_0w).$$

Therefore

$$
\begin{aligned}
J(B) &= \|z - ABw\|^2 \\
&= (z, z) - (z, ABw) - (ABw, z) + (ABw, ABw) \\
&= (z, z) - (AB_0w, ABw) - (ABw, AB_0w) \\
&\quad + (ABw, ABw) + (AB_0w, AB_0w) \\
&\quad - (AB_0, AB_0w) \\
&= \|A(B - B_0)w\|^2 + (z, z) - (AB_0w, AB_0w) \\
&= \|A(B - B_0)w\|^2 + (z, z) - (z, AB_0w) \\
&\quad + [(AB_0w, AB_0w) - (AB_0w, z)] \\
&= J(B_0) + \|A(B - B_0)w\| \geqslant J(B_0).
\end{aligned}
$$

Thus B_0 minimizes the given performance measured over $\mathcal{C}$ whenever it exists.

Remark 4.2. Theorem 4.1 reduces a very general optimization problem to an interpolation problem. We will show that in discrete time the interpolation problem is easily solved. We assume for now that $\mathcal{H} = l^2(0, \infty)$. This assumption is made purely for technical reasons and the results given here are true in general [5]. In our case, we can actually (by induction) construct the interpolating operator. We want to interpolate two vectors $\{x, y\}$ in $\mathcal{H}$ by means of an operator whose matrix representation w.r.t. the standard basis is lower triangular. It is clear that some condition is necessary. No lower triangular matrix will take the vector with first coordinate zero to a vector whose first coordinate is nonzero. More generally, no causal operator will take a vector in $E_n\mathcal{H}$ to a vector y for which $E^n y \neq 0$. This is the only restriction.

THEOREM 4.2. *Let $x, y \in l^2(0, \infty)$. There exists $A \in \mathcal{C}$ such that $Ax = y$ if and only if there exists a constant K such that for each $0 \leqslant r < \infty$*

$$\sum_{i=0}^{r} |y_i|^2 \leqslant K^2 \sum_{i=0}^{r} |x_i|^2,$$

or (*equivalently*)

$$\|E^r y\|^2 \leqslant K^2 \|E^r x\|^2.$$

Then A can be chosen with $\|A\| \leqslant K$.

Proof. The proof is by induction. We begin with the first row in A. Since $\|E^0 y\| \leqslant K\|E^0 x\|$, $E^0 x = 0$ implies $E^0 y = 0$. If $E^0 x = 0$, choose the first row in A to have all its entries zero. If $E^0 x = x_0 \neq 0$, choose a_{00} such that $y_0 = a_{00} x_0$ and $a_{0i} = 0$ for $i > 0$, since A must be lower triangular. Note that $|a_{00}| \leqslant K^2$ by hypothesis. So assume we have constructed the first $p-1$ rows of A such that if A_{p-1} is the $(p-1) \times (p-1)$ submatrix defined above, then $A_{p-1} E^{p-1} x = E^{p-1} y$ and $\|A_{p-1}\| \leqslant K$. We will now show how to construct the pth row. If a_{ij} is the i, jth element of A_{p-1}, then

$$y_i = \sum_{j=0}^{p-1} a_{ij} x_j, \qquad i = 0, 1, 2, \ldots p-1,$$

and $a_{ij} = 0$ for $j > i$. Also

$$\sum_{i=0}^{p-1} \left| \sum_{j=0}^{i} a_{ij} Z_j \right|^2 \leqslant K^2 \sum_{i=0}^{p-1} |z_i|^2,$$

where $(z_0, \ldots, z_{p-1}) = E^{p-1} z$, $z \in \mathcal{K}$.

Since $\|e^p y\|^2 \leqslant K^2 \|E^p x\|^2$, it follows that

$$\begin{aligned} |y_p|^2 &= \|E^p y\|^2 - \|E^{p-1} y\|^2 \\ &\leqslant K^2 \|E^p x\|^2 - \|E^{p-1} y\|^2 \\ &\leqslant K^2 \sum_{i=0}^{p-1} |x_i|^2 + K^2 |x_p|^2 - \sum_{i=0}^{p-1} \left| \sum_{j=1}^{i} a_{ij} x_j \right|^2. \end{aligned}$$

Define a linear functional on the one dimensional subspace in $E^p \mathcal{K}$ defined by $E^p x$ by

$$\phi(c E^p x) = c y_p,$$

and define a seminorm $\|_s$ on $E^p \mathcal{K}$ by

$$\|E^p z\|_s = K^2 |z_p|^2 + K^2 \sum_{i=0}^{p-1} |z_i|^2 - \sum_{i=0}^{p-1} \left| \sum_{j=0}^{i} a_{ij} z_j \right|^2.$$

Then $|\phi(c E^p x)| \leqslant \|E^p x\|_s$, and by the Hahn–Banach theorem ϕ can be extended to $E^p \mathcal{K}$ such that it satisfies the inequality

$$|\phi(E^p z)| \leqslant \|E^p z\|_s.$$

Since ϕ is a linear functional on $E^p \mathcal{K}$ there exist scalars $a_{p0}, \ldots, a_{pp}$ such that

$$\phi(E^p z) = \sum_{j=0}^{p} a_{pj} z_j.$$

These scalars will form the first p elements in the pth row with all other elements nonzero. Then

$$\begin{aligned}
\|A_p x\|^2 &= \sum_{i=0}^{p} |y_i|^2 \\
&= \sum_{i=0}^{p-1} |y_i|^2 + |y_p|^2 \\
&= \sum_{i=0}^{p-1} \left| \sum_{j=0}^{i} a_{ij} x_j \right|^2 + \left| \sum_{j=0}^{p} a_{pj} x_j \right|^2 \\
&= \sum_{i=0}^{p-1} \left| \sum_{j=0}^{i} a_{ij} x_j \right|^2 + |\phi(E^p x)|^2 \\
&\leqslant \sum_{i=0}^{p-1} \left| \sum_{j=0}^{i} a_{ij} x_j \right|^2 + K^2 |x_p|^2 + K^2 \sum_{i=0}^{p-1} |x_i|^2 \\
&\quad - \sum_{i=0}^{p-1} \left| \sum_{j=0}^{i} a_{ij} x_j \right|^2 \\
&= K^2 \|E^p x\|^2.
\end{aligned}$$

This completes the proof.

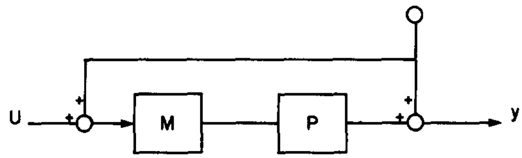
Remark. The Hahn–Banach theorem in the finite dimensional case is completely constructive, so there is no dependence on Zorn's Lemma here.

We apply this result to an elementary servo-mechanism problem [4]. Here we have an open loop system with plant P and compensator M that is disturbed at the output by a signal $n \in \mathcal{H}_t$. Our problem is to operate on n with a causal compensator that minimizes

$$J(M) = \|y\|^2 + \|r\|^2$$

over $M \in \mathcal{C}$.

Physically the requirement that $\|y\|^2$ is minimized means that the output of the plant must follow $-n$, so that the problem is really an optimal tracking problem. Of course, since P is linear, if one could apply arbitrarily large inputs r to P, this tracking problem would be straightforward. In practice, however, the norm of the inputs must be limited, and hence we minimize the sum of $\|y\|^2$ and $\|r\|^2$ to obtain a compromise between the tracking requirement and input energy.



THEOREM 4.3. *Suppose $P \in$ C. Then for y and r defined above and $n \in \mathcal{H}_t$,*

$$J(M) = \|y\|^2 + \|r\|^2$$

is minimized over $M \in \mathcal{C}$ by any $M_0 \in \mathcal{C}$ satisfying

$$M_0 n = -C^{-1} E_t C^{*-1} P^* n.$$

Proof. We reformulate the problem in terms of Theorem 4.1. Let $z = (n, 0) \in \mathcal{H} \oplus \mathcal{H}$, $w = n \in \mathcal{H}_t$, and let $B = M$ be causal on $\mathcal{H}$.

Define $A\colon \mathcal{H} \to \mathcal{H} \oplus \mathcal{H}$ by

$$A = \begin{pmatrix} -P \\ -I \end{pmatrix},$$

and let C^*C be the spectral factorization (by Corollary 2.2.) of $I + P^*P$. Then $z - ABw = \begin{pmatrix} n + PMn \\ Mn \end{pmatrix} = \begin{pmatrix} y \\ r \end{pmatrix}$. Thus minimizing $\|z - ABy\|^2$ over $B \in \mathcal{C}$ is equivalent to minimizing $\|y\|^2 + \|r\|^2$ over $M \in \mathcal{C}$. Since $A^*A = I + P^*P = C^*C$, we have

$$\begin{aligned} M_0 n = B_0 y &= C^{-1} E_t C^{*-1} A^* z \\ &= C^{-1} E_t C^{*-1} (-P^*, -I) \begin{pmatrix} n \\ 0 \end{pmatrix} \\ &= -C^1 E_t C^{*-1} P^* n, \end{aligned}$$

and the proof is complete.

5. STOCHASTIC CONTROL

In this section we will assume familiarity with the theory of Hilbert space-valued random variables. A detailed discussion can be found in [2], while the basic results necessary for following this section are given in [5]. We include here just the basic facts.

If X is a finite second moment $\mathcal{H}$-valued random variable (i.e., $E|(f, X(s))|^2$ is a continuous scalar valued mapping on $\mathcal{H}$), let $m_X \in \mathcal{H}$ denote its mean. Since $m_{X+Y} = m_X + m_Y$ [2], we can (by a suitable translation if necessary) assume that all random variables which we consider have zero mean. For any such random variables X, Y, let $Q_{XY} \in \mathcal{B}(\mathcal{H})$ be the cross-covariance of X and Y, and if $X = Y$ let $Q_X = Q_{XX}$ denote its co-variance. X and Y are said to be independent if $Q_{XY} = 0$. $A \in \mathcal{B}(\mathcal{H})$ is memoryless if $AE^n = E^nA$ for all $n \in Z$.

LEMMA 5.1. *Let X and Y be zero mean $\mathcal{H}$-valued random variables with finite second moment and $A, B \in \mathcal{B}(\mathcal{H})$. Then*

(i) $Q_{X+Y} = Q_X + Q_{XY} + Q_{YX} + Q_Y$,

(ii) $Q_{X+Y} = Q_X + Q_Y$ *if and only if X and Y are independent,*

(iii) $Q_{(AX)(BY)} = AQ_{XY}B^*$,

(iv) $Q_{AX} = AQ_XA^*$,

(v) $Q_{XY} = Q_{YX}^*$,

(vi) $Q_X = Q_X^* \geqslant 0$.

Proof. See [2, p. 249].

DEFINITION 5.2. A zero-mean $\mathcal{H}$-valued random variable x with finite second moment is *white noise* if $E^n x$ is independent of $E_n x$ for all $n \in Z$.

LEMMA 5.3. [5] *Let X be a zero mean $\mathcal{H}$-valued random variable with finite second moment. Then X is white noise if and only if Q_X is memoryless.*

Proof. If Q_X is memoryless, then

$$Q_{(E^nX)(E_nX)} = E^nQ_XE_n^* = E^nQ_XE_n = E^nE_nQ_X = 0.$$

On the other hand, if E^nX and E_nX are independent for all $n \in Z$,

$$E^nQ_XE_n = E^nQ_XE_n^* = Q_{(E^nX)(E_nX)} = 0,$$

and therefore $Q_X \in \mathcal{C}$. Similarly, $E_nQ_XE^n = 0$ and $Q_X \in \mathcal{C}^*$. Thus Q_X is memoryless. We can now state a basic stochastic optimization problem which is similar to the deterministic problem stated before.

A difficulty that arises in this setting is the choosing of a cost functional. In the classical theory of finite dimensional systems the cost functional that is chosen is

$$J(B) = E\|Z - ABW\|^2,$$

where B varies over $\mathcal{C}$, A is a bounded linear operator such that $A^*A > 0$, and Z, W are Hilbert space valued random variables. Unfortunately, in general, $J(B)$ may not be finite and in fact is finite if and only if its covariance operator Q_{Z-ABW} is trace class [2, 4]. This is quite a strong requirement and we can avoid it by considering a slightly more complicated cost-functional.

To motivate this, we first consider the finite dimensional case; Suppose X is an R^n-valued random variable, $X = (X_1, \cdots, X_n)$. Then

$$Q_X = EXX^t = \begin{bmatrix} EX_1^2 & EX_1X_2 & \cdots & EX_1X_n \\ EX_nX_1 & \cdots\cdots\cdots\cdots & & EX_n^2 \end{bmatrix},$$

and $E\|X\|^2 = \operatorname{tr}(Q_X) = \sum_{i=1}^n EX_i^2$. Thus the data for $E\|X\|^2$ are all contained on the diagonal of Q_X. We make use of the same data but in the form of a positive Hermitian operator

$$\mathcal{D}[Q_X] = \begin{bmatrix} EX_1^2 & & 0 \\ & \ddots & \\ 0 & & EX_n^2 \end{bmatrix},$$

and will minimize the cost functional in the ordering of positive Hermitian operators. This generalizes without difficulty to the infinite-dimensional discrete time case (in fact to the general infinite-dimensional case [7, 5]). It is easily seen that $\mathcal{D}[Q_X] = \sum_{i=-\infty}^{\infty} \Delta E_i Q_X \Delta E_i$ is a well-defined bounded linear operator on $\mathcal{H}$ and is memoryless. We list some simple properties of $\mathcal{D}[T]$ for $T \in \mathcal{B}(\mathcal{H})$ which we will make use of.

LEMMA 5.4. *The mapping $T \to \mathcal{D}(T)$ has the following properties*:

(i) *It is linear.*

(ii) *If T is Hermitian so is $\mathcal{D}[T]$.*

(iii) *If $T_1 > T_2$ then $\mathcal{D}[T_1] > \mathcal{D}[T_2]$.*

(iv) *If the diagonal elements of the matrix representation of T are zero then so is $\mathcal{D}[T]$.*

Proof. [7].

There is one more operation on the matrix representations of operators on $(\mathcal{H}, \mathcal{E})$ that we will need.

If $A \in \mathcal{B}(\mathcal{H})$, there is a natural (and unique) way to write A as the sum of a causal and anticausal operator. Define $[A]_{\mathcal{C}}$ by

$$\begin{aligned} ([A]_{\mathcal{C}})_{ij} &= (A)_{ij}, \qquad & i \geq j, \\ &= 0, & i < j. \end{aligned}$$

Then $[A]_{\mathcal{C}}$ is a lower triangular matrix, and let $[A]_{\mathcal{C}^*} = A - [A]_{\mathcal{C}}$. Unfortunately, $[A]_{\mathcal{C}}$ may not define a bounded operator on $\mathcal{H}$ as the following example shows [15, 5].

EXAMPLE 5.4. Let A be the semi-infinite Toeplitz matrix

$$A = \begin{bmatrix} 0 & -1 & -\frac{1}{2} & -\frac{1}{3} & -\frac{1}{4} & \cdots \\ 1 & 0 & -1 & -\frac{1}{2} & -\frac{1}{3} & \cdots \\ \frac{1}{2} & 1 & 0 & -1 & -\frac{1}{2} & \cdots \\ \frac{1}{3} & \frac{1}{2} & 1 & 0 & -1 & \cdots \\ \frac{1}{4} & \frac{1}{3} & \frac{1}{2} & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix},$$

defined by the sequence

$$\begin{aligned} a_{ik} = a_{i(i-k)} &= 1/k, \qquad & k \neq 0, \\ &= 0, & k = 0. \end{aligned}$$

Then $\|A\| = ess\sup|\hat{a}(\theta)|$, where $\hat{a}$ is the Fourier transform of the sequence $\{a_k\}$ defined on the unit disc. A simple computation shows that $\hat{a}(\theta) = 1/i(\pi - \theta)$ and $\|A\| = \pi$. But

$$[A]_{\mathcal{C}} = \begin{bmatrix} 0 & & & \bigcirc \\ 1 & 0 & & \\ \frac{1}{2} & 1 & \ddots & \\ \frac{1}{3} & \frac{1}{2} & \ddots & \\ \frac{1}{4} & & \ddots & \\ \vdots & & & \end{bmatrix}$$

is defined by the sequence

$$\begin{aligned} b_k &= \frac{1}{k}, \qquad & k > 0, \\ &= 0, & k \leqslant 0, \end{aligned}$$

and $\hat{b}(\theta) = -\ln(1 - e^{-i\theta})$, which is an unbounded function on the unit circle. Thus $[A]_{\mathcal{C}}$ is an unbounded operator on $\mathcal{H}$. In solving the optimal control problems it will be necessary to take the causal part of an operator, and we will assume that this exists as a bounded linear operator on $\mathcal{H}$.

THEOREM 5.5. [7] *Suppose $A \in \mathcal{C} \wedge \mathcal{C}^{-1}$ and Z, W are zero mean $\mathcal{H}$-valued random variables with finite second moment and covariance operators*

$Q_Z, Q_W > 0$. Let $DD^* = Q_W$ be the Arveson Factorization of Q_W, $D \in \mathcal{C} \cap \mathcal{C}^{-1}$. If $[Q_{ZW}D^{*-1}]_{\mathcal{C}}$ exists, then the cost functional

$$J(B) = \mathfrak{D}[Q_{Z-ABW}]$$

is minimized in the ordering of positive Hermitian operators over $B \in \mathcal{C}$ by

$$B_0 = A^{-1}\big[Q_{ZW}D^{*-1}\big]_{\mathcal{C}}D^{-1}.$$

Proof.

$$\begin{aligned} Q_{Z-ABW} &= Q_Z + Q_{(Z)(-ABW)} + Q_{(-ABW)(Z)} + Q_{ABW} \\ &= ABQ_W B^*A - ABQ_{WZ} - Q_{ZW}B^*A^* + Q_Z. \end{aligned}$$

Thus

$$\begin{aligned} \mathfrak{D}[Q_{Z-ABW}] &= \mathfrak{D}[ABQ_W B^*A^*] - \mathfrak{D}[ABQ_{WZ}] - \mathfrak{D}[Q_{ZW}B^*A^*] + \mathfrak{D}[Q_Z] \\ &= \mathfrak{D}[ABDD^*B^*A] - \mathfrak{D}\big[ABDD^{-1}Q_{WZ}\big] \\ &\quad - \mathfrak{D}\big[Q_{ZW}D^{-1*}D^*B^*A^*\big] + \mathfrak{D}[Q_Z]. \end{aligned}$$

Let $X = Q_{ZW}D^{*-1}$, and rewrite this as

$$\begin{aligned} &\mathfrak{D}[ABDD^*B^*A] - \mathfrak{D}[ABDX^*] - \mathfrak{D}[XD^*B^*C^*] + \mathfrak{D}[Q_Z] \\ &\qquad = \mathfrak{D}[ABDD^*B^*A^* - ABDX^* - XD^*B^*A^*] + \mathfrak{D}[Q_Z] \\ &\qquad = \mathfrak{D}[(ABD - X)(ABD - X)^*] + \mathfrak{D}[Q_Z - XX^*]. \end{aligned}$$

Since $Q_Z - XX^*$ is independent of B, and $\mathfrak{D}$ is positive, a minimizing $B \in \mathcal{B}(\mathcal{H})$ can be obtained from the above expression by letting $B = A^{-1}XD^{-1}$, since for this B we obtain $J(B) = \mathfrak{D}[Q_Z - XX^*]$, which is independent of B, and $\mathfrak{D}[(ABD - X)(ABD - X)^*] \geqslant 0$.

Unfortunately, $B \notin C$ in general. So let $X = [X]_{\mathcal{C}} + [X]_{\mathcal{C}^*}$, and substitute this into the cost functional. Then

$$\begin{aligned} J(B) &= \mathfrak{D}[(ABD - [X]_{\mathcal{C}} - [X]_{\mathcal{C}^*}) \\ &\qquad \times (ABD - [X]_{\mathcal{C}} - [X]_{\mathcal{C}^*})^*] + \mathfrak{D}[Q_Z - XX^*] \\ &= \mathfrak{D}[(ABD - [X]_{\mathcal{C}})(ABD - [X]_{\mathcal{C}})^*] \\ &\qquad - \mathfrak{D}[(ABD - [X]_{\mathcal{C}})[X]_{\mathcal{C}^*}^*] - \mathfrak{D}[[X]_{\mathcal{C}^*}(ABD - [X]_{\mathcal{C}})^*] \\ &\qquad + \mathfrak{D}[Q_Z - XX^* + [X]_{\mathcal{C}^*}[X]_{\mathcal{C}^*}^*]. \end{aligned}$$

Now $ABD - [X]_{\mathcal{C}}$ is lower triangular, and $[X]_{\mathcal{C}^*}^*$ is strictly lower triangular. Thus so is their product, and $\mathfrak{D}[(ABD - [X]_{\mathcal{C}})[X]_{\mathcal{C}^*}^*] = 0$. By the same reasoning $\mathfrak{D}[[X]_{\mathcal{C}*}[ABD - [X]_{\mathcal{C}})^*] = 0$, and

$$J(B) = \mathfrak{D}[(ABD - [X]_{\mathcal{C}})(ABD - [X]_{\mathcal{C}})^*] + \mathfrak{D}[Q_Z - XX^* + [X]_{\mathcal{C}*}[X]^*_{\mathcal{C}*}].$$

This is minimized by

$$B_0 = A^{-1}[X]_{\mathcal{C}} D^{-1},$$

since $\mathfrak{D}[(ABD - [X]_{\mathcal{C}})(ABD - [X]_{\mathcal{C}})^*] \geqslant 0$, and $\mathfrak{D}[Q_Z - XX^* + [X]_{\mathcal{C}*} [X]^*_{\mathcal{C}*}]$ is independent of B.

Remarks.

(1) The price we pay for a causal solution is

$$\mathfrak{D}[[X]_{\mathcal{C}*}[X]^*_{\mathcal{C}*}].$$

(2) The assumption that $Q_W > 0$, is not really restrictive. If W contains white noise with identity covariance, then Q_W is positive definite. Since this is the only assumption we made, it becomes clear that Theorem 5.5 is quite general.

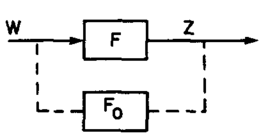
6. Applications

In this section we apply Theorem 5.5 to solve some general problems in stochastic systems.

System Identification

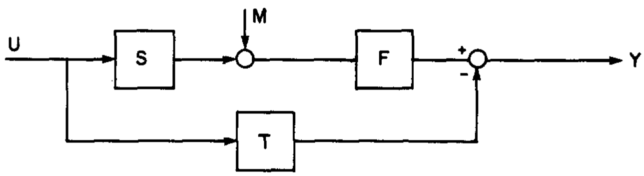
We observe the input and output processes W and Z respectively associated with an unknown system. Since the system is "physical," we may assume that it is causal and thus identify it by determining $F \in \mathcal{C}$ that minimizes $\mathfrak{D}[Q_{Z-FW}]$. This is just a special case of Theorem 5.5 with $A = I$, if we assume $Q_W > 0$. If $Q_W = DD^*$, then the solution is given by

$$F_0 = [Q_{ZW} D^{*-1}]_{\mathcal{C}} D^{-1}.$$



Wiener Filter and Predictor

We consider the following configuration. U is a signal process of some type that is observed through a sensor S, whose output is corrupted by a noise term M. On the basis of these observations, we desire to construct an optimal causal Filter F that processes these observations to obtain an estimate of TU, where $T \in \mathcal{B}(\mathcal{H})$ is a given linear operator. $T = I$ is the case of the linear filter, and $T = P$ the ideal predictor of $L^2(-\infty, \infty)$ is the Wiener predictor.



We assume U and M are zero mean H-valued random variables with finite second moment and covariances Q_U, Q_M with $Q_M > 0$. Let DD^* be the spectral factorization of $[Q_M + SQ_{UM} + Q_{MU}S^* + SQ_US^*]$. Then $\mathcal{D}[Q_Y]$ is minimized over $F \in \mathcal{C}$ by

$$F_0 = \left[T(Q_{UM} + Q_US^*)D^{*-1}\right]_{\mathcal{C}}D^{-1}.$$

Proof. Apply Theorem 5.5 with $Z = TU$, $W = SU + M$. Then $Q_Z = TQ_UT^*$, and $Q_W = Q_M + SQ_{UM} + Q_{MU}S^* + SQ_US^* = DD^* > 0$, since $Q_M > 0$. Finally

$$Q_{ZW} = Q_{(TU)(SU+M)} = TQ_US^* + TQ_{UM},$$

and thus $F_0 = [T(Q_{UM} + Q_US^*)D^{*-1}]_{\mathcal{C}}D^{-1}$.

References

1. W. A. Arveson, Interpolation in nest algebras, *J. Funct. Anal.* **20** (1975), 208–233.
2. H. V. Balakrishnan, "Applied Functional Analysis," Springer-Verlag, Berlin New York, 1976.
3. M. Choi and A. Feintuch, Operator factorization with respect to nest algebras, preprint, 1981.
4. R. M. DeSantis, R. Saeks, and L. J. Tung, Basic optimal estimation and control problems in Hilbert space, *Math. Systems Theory* **12** (1978), 175–203.
5. A. Feintuch and R. Saeks, "System Theory: A Hilbert Space Approach," Academic Press, New York, 1982.

6. A. FEINTUCH AND R. SAEKS, Extended spaces and the resolution topology, *Internat. J. Control* **39** (1981), 347–354.
7. A. FEINTUCH, R. SAEKS, AND C. NEIL, A new performance measure for stochastic optimization in Hilbert space, *Math. Systems Theory* **15** (1981), 39–54.
8. A. FEINTUCH, Strong causality conditions and causal invertibility, *SIAM J. Control Optim.* **18** (1980), 317–324.
9. R. B. HOLMES, Mathematical Foundations of Signal Processing, *SIAM Rev.* **21** (1979), 361–389.
10. J. HORVATH, "Topological Vector Spaces," Addison–Wesley, Reading, Mass., 1966.
11. D. LARSON, Nest algebras and similarity transformations, preprint, Univ. of Nebraska, 1982.
12. W. A. PORTER, A basic optimization problem in linear systems, *Math. Systems Theory* **5** (1971), 20–44.
13. W. A. PORTER AND C. L. ZAHM, "Basic Concepts in System Theory," Tech. Rep. 33, Systems Eng. Lab., Univ. of Michigan, 1969.
14. R. SAEKS, Causality in Hilbert space, *SIAM Rev.* **12** (1970), 357–383.
15. R. SAEKS, R. M. DESANTIS, AND R. J. LEAKE, On causal decomposition, *IEEE Trans. Autom. Control* **19** (1974), 152–153.
16. R. SAEKS, The factorization problem: A survey, *IEEE Proc.* **64** (1976), 90–95.
17. A. SCHUMITZKY, State space control for general linear systems, *Proc. Int. Symp. MTNS* **3** (1979), 194–204.
18. J. RISSANEN AND L. BARBOSA, A factorization problem and the problem of predicting non-stationary vector-valued stochastic processes, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **12** (1969), 255–266.
19. J. RISSANEN, On factoring positive operators, *J. Math. Anal. Appl.* **32** (1970), 505–511.